On Functions of Companion Matrices

József Eller
Computing Centre
Szeged University of Medicine
Szeged, Hungary

Submitted by R. E. Hartwig

ABSTRACT

Presented in this paper are some new properties of a function $f(C)$ of a companion matrix C, including (1) a representation of any entry of $f(C)$ as a divided difference of $f(\lambda)$ times a polynomial, (2) a similarity decomposition of $f(C)$ generalizing that based on the Jordan form of C, and (3) characterization (construction) of all matrices that transform $f(C)$ (by similarity) to companion form. The connection between functions of general and companion matrices is also dealt with, and a pair of dual relations is established.

1. INTRODUCTION

Companion matrices are often encountered in linear algebra and applied mathematics (e.g., Frobenius normal form of matrices, phase-variable canonical form of dynamical systems). They are a common device for transforming polynomial root problems into eigenvalue problems, or higher-order linear differential equations into first-order linear systems. Elementary properties of companion matrices were summarized by Brand [2]. Some useful results for handling companion matrices are contained in Hartwig [8, 9] and Young [17].

In this paper we deal with various representations related to the matrix function $f(C)$ of a companion matrix C. These representations generally involve divided differences of the function $f(\lambda)$, the power functions as well as the adjoint polynomials associated with C, evaluated at the eigenvalues of C. Thus, in Section 2 we derive such a representation for the entries of $f(C)$, where the power and adjoint polynomial involved depend on the particular entry.

Section 3 deals with similarity transformation of $f(C)$ to block diagonal form, each block being associated with a cluster of eigenvalues of C. If the eigenvalues within clusters coincide, the well-known representation of $f(C)$ based on Jordan form of C results. The divided-difference representations of the similarity transformation matrix (generalized Vandermonde matrix) and that of its inverse are also determined. As a consequence the similarity transformation of C to the direct sum of smaller companion matrices is obtained, too.

In Section 4 we give a characterization of all similarity transformation matrices that take $f(C)$ to companion form, obtaining as a special case the result of Shane and Barnett for linear fractional functions ([15]; see also [17]). Divided-difference representations of the transformation matrices and their inverses are given, too.

In the last section we generalize a theorem of Stafney [16] stating that for a general square matrix A, $f(A)$ is equal to a polynomial in A, with the entries of the first row of $f(C)$ as coefficients, C being the companion matrix of the minimal polynomial of A. We also obtain a dual relationship, where powers of A are replaced by adjoint polynomials of A.

Throughout the paper, column vectors are denoted by boldface lowercase letters; in particular, $\mathbf{e}_1 = [1,0,\dots,0]^T,\dots,\mathbf{e}_n = [0,\dots,0,1]^T$ denote the n natural unit vectors. The equality $A = (a_{ij})_{m\times n}$ defines a real or complex $m \times n$ matrix with its (i, j) entry given by a_{ij}. In particular, $I = (\delta_{ij})_{n\times n}$ defines the identity matrix, where δ_{ij} is the Kronecker delta. The companion matrix of the polynomial $p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n$ (λ a complex variable) is defined as

$$C = \left(\delta_{i,j-1} - c_{j-1}\delta_{in}\right)_{n\times n}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \cdot & 0 & 1 & 0 & \cdots & 0 \\ \cdot & & 0 & 1 & \ddots & \vdots \\ \cdot & & & \ddots & \ddots & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{n-2} & -c_{n-1} \end{bmatrix}, \tag{1.1}$$

while the adjoint polynomials of $p(\lambda)$ are given by

$$\begin{aligned} p_{n-1}(\lambda) &= c_n = 1, \\ p_{j-1}(\lambda) &= c_j + \lambda p_j(\lambda) \\ &= c_j + c_{j+1}\lambda + \cdots + c_n\lambda^{n-j}, \qquad j = n-1,\dots,2,1. \end{aligned} \tag{1.2}$$

The kth-order divided difference of a scalar function $f(\lambda)$ at the (not necessarily distinct) points $\lambda_0, \lambda_1, \dots, \lambda_k$ will be denoted by $f[\lambda_0; \lambda_1; \dots; \lambda_k]$, and the kth derivative function by $f^{(k)}(\lambda)$.

When dealing with a matrix function $f(A)$ with some square matrix A and scalar function $f(\lambda)$ *we shall always assume* (without reiterating) that $f(\lambda)$ is analytic on a domain of the complex plane containing the eigenvalues of A. As our results involve matrix functions with fixed arguments only, this assumption does not really restrict generality, since, by definition [7], $f(A)$ equals a polynomial in A, which is even an entire function. In particular, $f(A) = h(A)$, where $h(\lambda)$ is the Hermite interpolation polynomial of $f(\lambda)$ at the eigenvalues of A, with multiplicities determined by the minimal polynomial of A. Also, the divided differences of $f(\lambda)$ and $h(\lambda)$ agree at any subset of the roots of the minimal polynomial of A.

Under the previous assumption we can use the contour-integral representation of $f(A)$ given by

$$f(A) = \frac{1}{2\pi i} \int_\gamma f(\lambda)(\lambda I - A)^{-1} d\lambda \tag{1.3}$$

where γ is a simple closed curve lying in the domain of analyticity of $f(\lambda)$ that encircles all eigenvalues of the matrix A [14]. We shall also use the counterpart of (1.3) for divided differences, i.e., the representation

$$f[\lambda_0; \lambda_1; \dots; \lambda_k] = \frac{1}{2\pi i} \int_\gamma \frac{f(\lambda)}{\prod_{j=0}^{k}(\lambda - \lambda_j)} d\lambda, \tag{1.4}$$

where the points $\lambda_0, \lambda_1, \dots, \lambda_k$ are contained in the interior of the contour γ [1, p. 172].

2. DIVIDED-DIFFERENCE REPRESENTATION OF THE ELEMENTS OF $f(C)$

For establishing the divided-difference representation of the entries of the matrix function $f(C)$ of the $n \times n$ companion matrix C, we use a lemma giving a dual pair of relations which follow simply from the row-column structure of C. We assume in this section that the companion matrix C and the adjoints $p_j(\lambda)$ of the polynomial $p(\lambda)$ are defined by (1.1) and (1.2), respectively.

LEMMA 2.1. *The companion matrix C satisfies the relations*

$$\mathbf{e}_1^T C^{i-1} = \mathbf{e}_i^T, \qquad i = 1, 2, \ldots, n, \tag{2.1}$$

$$p_{j-1}(C)\mathbf{e}_n = \mathbf{e}_j, \qquad j = n, n-1, \ldots, 1. \tag{2.2}$$

Proof. Both relations follow easily by induction, making use of the relations $\mathbf{e}_{i-1}^T C = \mathbf{e}_i^T$ and $C\mathbf{e}_j = \mathbf{e}_{j-1} - c_{j-1}\mathbf{e}_n$ $(i, j = 1, \ldots, n-1;\ \mathbf{e}_0 = \mathbf{0})$, as well as the recursion in (1.2). ∎

Now we can derive the following representation for $f(C)$.

THEOREM 2.1. *The elements of $f(C) = (f_{ij})_{n \times n}$ can be represented as*

$$f_{ij} = \mathbf{e}_1^T C^{i-1} f(C) p_{j-1}(C)\mathbf{e}_n = \varphi_{ij}[\lambda_1; \ldots; \lambda_n], \qquad i, j = 1, \ldots, n, \tag{2.3}$$

where the functions $\varphi_{ij}(\lambda)$ are defined by

$$\varphi_{ij}(\lambda) = \lambda^{i-1} f(\lambda) p_{j-1}(\lambda), \tag{2.4}$$

and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of C.

Proof. Using Lemma 2.1 and the relations (1.3)–(1.4), we have

$$\begin{aligned} f_{ij} &= \mathbf{e}_i^T f(C)\mathbf{e}_j = \mathbf{e}_1^T C^{i-1} f(C) p_{j-1}(C)\mathbf{e}_n = \mathbf{e}_1^T \varphi_{ij}(C)\mathbf{e}_n \\ &= \frac{1}{2\pi i}\int_\gamma \varphi_{ij}(\lambda)\mathbf{e}_1^T(\lambda I - C)^{-1}\mathbf{e}_n\, d\lambda = \frac{1}{2\pi i}\int_\gamma \frac{\varphi_{ij}(\lambda)}{p(\lambda)}\, d\lambda \\ &= \varphi_{ij}[\lambda_1; \ldots; \lambda_n], \qquad i, j = 1, \ldots, n, \end{aligned}$$

where we have used the fact that the cofactor of the $(1, n)$ entry in the determinant of $\lambda I - C$ was equal to 1, and $\det(\lambda I - C) = p(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i)$. ∎

Theorem 2.1 can be employed to bound elements of $f(C)$ by means of bounding divided differences. For this purpose, the following quadrature representation of divided differences is useful [1, p. 188]:

$$f[\lambda_1; \ldots; \lambda_n] = \int \cdots \int f^{(n-1)}(\lambda_1 x_1 + \cdots + \lambda_n x_n)\, dx_1 \cdots dx_{n-1}, \tag{2.5}$$

where the integral extends over all positive values of $x_1,\dots,x_{n-1}$ that satisfy $1-x_n=x_1+\cdots+x_{n-1}\leqslant 1$. For instance, in case of the $(1,n)$ entry of the matrix exponential $\exp(tC)$ we have $\mathbf{e}_1^T\exp(tC)\mathbf{e}_n=f[\lambda_1;\dots;\lambda_n]$, where $f(\lambda)=\exp(t\lambda)$; hence

$$|\mathbf{e}_1^T\exp(tC)\mathbf{e}_n|=\left|\int\cdots\int t^{n-1}\exp(\lambda_1x_1+\cdots+\lambda_nx_n)\,dx_1\cdots dx_{n-1}\right|$$

$$\leqslant\frac{t^{n-1}}{(n-1)!}\exp\Bigl(t\max_{1\leqslant i\leqslant n}\operatorname{Re}\lambda_i\Bigr),\qquad t\geqslant 0.$$

This bound was used in [5] to derive a perturbation bound for the solution of a linear nth-order differential equation with constant coefficients.

The following corollary to Theorem 2.1 will be needed in the next section. It follows by taking $f(\lambda)\equiv 1$ and noting that $f(C)=I$.

Corollary 2.1. *The divided differences of the functions*

$$\psi_{ij}(\lambda)=\lambda^{i-1}p_{j-1}(\lambda),\qquad i,j=1,\dots,n,$$

taken at the zeros $\lambda_1,\dots,\lambda_n$ of the polynomial $p(\lambda)$, are given by

$$\psi_{ij}[\lambda_1;\dots;\lambda_n]=\delta_{ij}.$$

3. A GENERALIZATION OF THE JORDAN REPRESENTATION OF $f(C)$

If the Jordan canonical decomposition of the companion matrix C is given by

$$C=VJ_CV^{-1},\qquad J_C=J_{n_1}(\lambda^{(1)})\oplus\cdots\oplus J_{n_s}(\lambda^{(s)}),\tag{3.1}$$

where $J_{n_k}(\lambda^{(k)})$ is an $n_k\times n_k$ Jordan matrix corresponding to the eigenvalue $\lambda^{(k)}$, then

$$f(C)=Vf(J_C)V^{-1},\qquad f(J_C)=f\bigl(J_{n_1}(\lambda^{(1)})\bigr)\oplus\cdots\oplus f\bigl(J_{n_s}(\lambda^{(s)})\bigr)\tag{3.2}$$

(cf. Corollary 3.1). Since a companion matrix is always nonderogatory, different Jordan blocks correspond to different eigenvalues. It is known [2] that the matrix V, having the principal vectors of C as its columns, can be chosen as the (possibly confluent) Vandermonde matrix associated with the eigenvalues of C.

Here we generalize the above relation so that C is similar to a block diagonal matrix $[\Lambda^{(1)}] \oplus \cdots \oplus [\Lambda^{(s)}]$, each block being associated with a cluster of eigenvalues of C, where different clusters have no eigenvalue in common. The similarity transformation matrix and its inverse will be given explicitly, in terms of divided differences of powers and rational functions (respectively). The blocks $[\Lambda^{(k)}]$ have a special form that makes possible to build up $f([\Lambda^{(k)}])$ from divided differences of $f(\lambda)$, as given in the following lemma.

LEMMA 3.1. *Let the matrices* $[\Lambda]$ *and* $f[\Lambda]$ *associated with the (possibly confluent) numbers* $\lambda_1, \dots, \lambda_n$ *be defined by*

$$[\Lambda] = \begin{bmatrix} \lambda_1 & 1 & & & 0 \\ & \lambda_2 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{bmatrix}, \tag{3.3}$$

$$f[\Lambda] = \begin{bmatrix} f[\lambda_1] & f[\lambda_1;\lambda_2] & \cdots & f[\lambda_1;\dots;\lambda_n] \\ & f[\lambda_2] & \cdots & f[\lambda_2;\dots;\lambda_n] \\ & & \ddots & \vdots \\ 0 & & & f[\lambda_n] \end{bmatrix}.$$

Then

$$f([\Lambda]) = f[\Lambda]. \tag{3.4}$$

Proof. Follows directly from Theorem 1.1 in Stafney [16] for upper triangular matrices. As Stafney's theorem concerns polynomials, it only is to be noted that $f(\lambda)$ can be replaced by its Hermite interpolation polynomial in (3.3)–(3.4). ∎

The statement of Lemma 3.1 was proved in the distinct-eigenvalue case by Opitz [13], who proposed the relation (3.3)–(3.4) for extending the definition of divided differences to the case of confluent arguments.

For the proof of our next theorem we need, in addition, the following useful result of Hartwig [9], which will also be employed later.

LEMMA 3.2 (Folklore lemma [9]). *Let C be the companion matrix of the polynomial $p(\lambda)$ of degree n, and A an $n \times n$ matrix. Then:*

(a) *The general solution of the matrix equation $AX = XC^T$ has the form $X = [\mathrm{y}, A\mathrm{y},\dots, A^{n-1}\mathrm{y}]$, where $p(A)\mathrm{y} = \mathbf{0}$.*

(b) *X is nonsingular if and only if A is nonderogatory and the minimal polynomial of the vector y is $p(\lambda)$.*

(c) *If $A = q(\tilde{C})$, and $\tilde{C}$ is the companion matrix of $\tilde{p}(\lambda)$, where $q(\lambda)$, $\tilde{p}(\lambda)$ are polynomials, then X is nonsingular if and only if the polynomials $y(\lambda) = y_1 + y_2\lambda + \cdots + y_n\lambda^{n-1}$ and $\tilde{p}(\lambda)$ are relative primes.*

Now we can establish the main result of this section.

THEOREM 3.1. *Let C be the companion matrix of the polynomial*

$$p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n = \prod_{k=1}^{s}\prod_{\nu=1}^{n_k}\left(\lambda - \lambda_\nu^{(k)}\right),$$

where $n = \sum_{k=1}^{s} n_k$ and $\lambda_\nu^{(k)} \neq \lambda_\mu^{(l)}$ for $k \neq l$. Then

$$f(C) = Q\left(f[\Lambda^{(1)}] \oplus \cdots \oplus f[\Lambda^{(s)}]\right)Q^{-1}, \tag{3.5}$$

where $[\Lambda^{(k)}] = (\lambda_i^{(k)}\delta_{ij} + \delta_{i,j-1})_{n_k\times n_k}$, $f[\Lambda^{(k)}]$ is defined according to (3.3), *and the matrix Q and its inverse are given in partitioned form as*

$$Q = \left[Q^{(1)},\dots,Q^{(s)}\right], \qquad Q^{-1} = \begin{bmatrix} \tilde{Q}^{(1)} \\ \vdots \\ \tilde{Q}^{(s)} \end{bmatrix} \tag{3.6a}$$

with

$$\left.\begin{aligned} Q^{(k)} &= \left(q_{ij}^{(k)}\right)_{n\times n_k}, & q_{ij}^{(k)} &= q_{i-1}\left[\lambda_1^{(k)};\dots;\lambda_j^{(k)}\right], \\ \tilde{Q}^{(k)} &= \left(\tilde{q}_{ij}^{(k)}\right)_{n_k\times n}, & \tilde{q}_{ij}^{(k)} &= p_{k,j-1}\left[\lambda_i^{(k)};\dots;\lambda_{n_k}^{(k)}\right], \end{aligned}\right\} \quad k = 1,\dots,s, \tag{3.6b}$$

with $q_{i-1}(\lambda)$ *and* $p_{k,j-1}(\lambda)$ *denoting the functions*

$$q_{i-1}(\lambda)=\lambda^{i-1},\qquad p_{k,j-1}(\lambda)=\frac{p_{j-1}(\lambda)}{\Pi_k(\lambda)},\tag{3.7}$$

where

$$\Pi_k(\lambda)=\frac{p(\lambda)}{\prod_{\nu=1}^{n_k}\left(\lambda-\lambda_\nu^{(k)}\right)}.$$

Proof. It suffices to prove that the two matrix equations

$$CQ=Q([\Lambda^{(1)}]\oplus\cdots\oplus[\Lambda^{(s)}]),\tag{3.8}$$

$$Q\tilde{Q}=I\tag{3.9}$$

hold, where $\tilde{Q}$ is the matrix given in (3.6) for Q^{-1}, since then the general case immediately follows by Lemma 3.1.

Applying the folklore lemma to (3.8) with $X=Q^T$, $A=([\Lambda^{(1)}]\oplus\cdots\oplus[\Lambda^{(s)}])^T$, and $\mathbf{y}^T=[\mathbf{e}_1^T,\dots,\mathbf{e}_1^T]$ (with consistent partition), we get that Equation (3.8) holds for

$$Q^{(k)}=\left[\mathbf{e}_1,[\Lambda^{(k)}]^T\mathbf{e}_1,\dots,\left([\Lambda^{(k)}]^{n-1}\right)^T\mathbf{e}_1\right]^T,$$

since $p(A)=0$. Now by Lemma 3.1 we obtain the desired divided-difference representation for the entries of $Q^{(k)}$:

$$q_{ij}^{(k)}=\mathbf{e}_1^T[\Lambda^{(k)}]^{i-1}\mathbf{e}_j=\mathbf{e}_1^Tq_{i-1}[\Lambda^{(k)}]\mathbf{e}_j=q_{i-1}\left[\lambda_1^{(k)};\dots;\lambda_j^{(k)}\right].\tag{3.10}$$

For the proof of (3.9) we first note that by the assumption of the theorem on the eigenvalues of C, there exist separate simple closed contours $\gamma_1,\dots,\gamma_s$ in the complex plane such that $\lambda_1^{(k)},\dots,\lambda_{n_k}^{(k)}$ lie inside γ_k $(k=1,\dots,s)$. It is also clear that the rational functions $p_{k,j-1}(\lambda)$ $(j=1,\dots,n)$ are analytic inside γ_k. Let γ be a contour encircling all eigenvalues of C. Then, using the Leibniz product rule for divided differences, the contour-integral representation (1.4), the residue theorem (see, e.g., [3]), and Corollary 2.1, we can proceed as follows:

$$\begin{aligned}
\mathbf{e}_i^T Q\tilde{Q}\mathbf{e}_j &= \sum_{k=1}^{s}\sum_{\nu=1}^{n_k} q_{i-1}\left[\lambda_1^{(k)};\ldots;\lambda_\nu^{(k)}\right] p_{k,j-1}\left[\lambda_\nu^{(k)};\ldots;\lambda_{n_k}^{(k)}\right]\\
&= \sum_{k=1}^{s}\left(q_{i-1}\times p_{k,j-1}\right)\left[\lambda_1^{(k)};\ldots;\lambda_{n_k}^{(k)}\right]\\
&= \sum_{k=1}^{s}\frac{1}{2\pi i}\int_{\gamma_k}\frac{q_{i-1}(\lambda)p_{k,j-1}(\lambda)}{\prod_{\nu=1}^{n_k}\left(\lambda-\lambda_\nu^{(k)}\right)}\,d\lambda\\
&= \frac{1}{2\pi i}\sum_{k=1}^{s}\int_{\gamma_k}\frac{q_{i-1}(\lambda)p_{j-1}(\lambda)}{p(\lambda)}\,d\lambda\\
&= \sum_{k=1}^{s}\operatorname*{Res}_{\gamma_k}\frac{q_{i-1}(\lambda)p_{j-1}(\lambda)}{p(\lambda)} = \operatorname*{Res}_{\gamma}\frac{q_{i-1}(\lambda)p_{j-1}(\lambda)}{p(\lambda)}\\
&= \frac{1}{2\pi i}\int_{\gamma}\frac{q_{i-1}(\lambda)p_{j-1}(\lambda)}{p(\lambda)}\,d\lambda\\
&= \left(q_{i-1}\times p_{j-1}\right)\left[\lambda_1^{(1)};\ldots;\lambda_{n_1}^{(1)};\ldots;\lambda_1^{(s)};\ldots;\lambda_{n_s}^{(s)}\right] = \delta_{ij},
\end{aligned}$$

where we use the notation $\operatorname{Res}_\gamma g(\lambda)$ for the sum of residues of the function $g(\lambda)$ at the poles inside the contour γ. ∎

The following corollary concerns the special case when the eigenvalues within clusters coincide, which yields the Jordan representation of $f(C)$. In this case, as easily seen from the formula (1.4), divided differences within clusters reduce to derivatives divided by factorials (according to Cauchy's contour-integral representation of derivatives). Correspondingly, the transformation matrix Q reduces to the confluent Vandermonde matrix V, the inverse of which is also obtained. (For further representations of the inverse of a confluent Vandermonde matrix see, e.g., [4].)

COROLLARY 3.1. *Let c be the companion matrix of the polynomial*

$$p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} = \prod_{k=1}^{s}\left(\lambda - \lambda^{(k)}\right)^{n_k},$$

where $\lambda^{(1)}, \ldots, \lambda^{(s)}$ *are distinct numbers. Then* $f(C)$ *is given by the relations* (3.1)–(3.2), *where*

$$J_{n_k}(\lambda^{(k)}) = \begin{bmatrix} \lambda^{(k)} & 1 & & & 0 \\ & \lambda^{(k)} & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & \lambda^{(k)} \end{bmatrix},$$

$$f\big(J_{n_k}(\lambda^{(k)})\big) = \begin{bmatrix} f(\lambda^{(k)}) & \dfrac{f'(\lambda^{(k)})}{1!} & \cdots & \dfrac{f^{(n_k-1)}(\lambda^{(k)})}{(n_k-1)!} \\ & f(\lambda^{(k)}) & \ddots & \vdots \\ & & \ddots & \dfrac{f'(\lambda^{(k)})}{1!} \\ 0 & & & f(\lambda^{(k)}) \end{bmatrix},$$

$$V = \left[V^{(1)}, \ldots, V^{(s)}\right],$$

$$V^{-1} = \begin{bmatrix} \tilde{V}^{(1)} \\ \vdots \\ \tilde{V}^{(s)} \end{bmatrix},$$

$$V^{(k)} = \left(\binom{i-1}{j-1} (\lambda^{(k)})^{i-j} \right)_{n_k \times n_k},$$

$$\tilde{V}^{(k)} = \left(\tilde{v}_{ij}^{(k)} \right)_{n_k \times n},$$

$$\tilde{v}_{ij}^{(k)} = \frac{1}{(n_k - i)!} p_{k,j-1}^{(n_k - i)}(\lambda^{(k)}),$$

where[1]

$$p_{k,j-1}(\lambda) = \frac{p_{j-1}(\lambda)}{p(\lambda)} (\lambda - \lambda^{(k)})^{n_k}$$

and $p_{j-1}(\lambda)$ *is the* $(j-1)$*st adjoint polynomial of* $p(\lambda)$.

---

[1]According to the extended definition of binomial coefficients, $\binom{i-1}{j-1} = 0$ if $j < 1$ or $j > i$.

REMARK 3.1. The decomposition of $f(C)$ [or that of C, if $f(\lambda)\equiv\lambda$] given by Theorem 3.1 has the advantage over that based on the Jordan form of C (see Corollary 3.1) that—assuming a fixed partition $\{\lambda_\nu^{(k)}\}$—it depends continuously on the eigenvalues of C [or equivalently, on the coefficients of $p(\lambda)$]. For, by the continuity of eigenvalues, small enough perturbations in the coefficients of $p(\lambda)$ keep the eigenvalue cluster $\{\lambda_\nu^{(k)},\ \nu=1,\dots,n_k\}$ inside the contour γ_k used in the proof of Theorem 3.1, and for an analytic $f(\lambda)$, all divided differences occurring in the decomposition (3.5)–(3.6) of $f(C)$ depend continuously on eigenvalues. On the other hand, arbitrarily small changes in the coefficients or eigenvalues may alter the whole structure of the Jordan based decomposition (3.1)–(3.2) in the confluent case, since some of the coincident eigenvalues may become distinct. [Consider, e.g., the polynomial $p(\lambda)=\lambda^2$ and its perturbation $\tilde{p}(\lambda)=\lambda^2+\varepsilon$, where the number ε can be arbitrarily small.] In general, an arbitrarily small nonzero perturbation in the confluent case will yield distinct, but nearly confluent eigenvalues and a corresponding nearly singular (ill-conditioned) Vandermonde transformation matrix.

REMARK 3.2. The divided-difference representations of $f(C)$ are not recommended for practical computations, as divided differences at nearly confluent arguments are known to be particularly sensitive to rounding error [12, p. 824]. Nevertheless, Theorem 3.1 provides the possibility of clustering nearby eigenvalues and handling the blocks $Q^{(k)}, \tilde{Q}^{(k)}$ and $f[\Lambda^{(k)}]$ by special techniques that avoid divided differences. For example, the rows of $Q^{(k)}$ can be obtained by successive vector-matrix multiplications according to (3.10). Or, in the case of the exponential function, the procedure of Evans et al. [6] can be mentioned, who compute divided differences at nearby arguments as upper triangular entries of the matrix function $\exp[\Lambda]$ according to Lemma 3.1. A partition into well-separated clusters should lead to a satisfactorily conditioned transformation matrix Q. Regarding the computation of the exponential of a general matrix A, such an approach was suggested as a most promising one [12, p. 825].

Another particular case of interest in Theorem 3.1 is when we have only one eigenvalue cluster ($s=1$). This case is formulated in the following corollary.

COROLLARY 3.2. *Let C be the companion matrix of*

$$p(\lambda)=c_0+c_1\lambda+\cdots+c_{n-1}\lambda^{n-1}+\lambda^n=\prod_{i=1}^{n}(\lambda-\lambda_i).$$

Then

$$f(C)=Rf[\Lambda]R^{-1},$$

where $[\Lambda]$ *and* $f[\Lambda]$ *are given by* (3.3) *and the elements of* $R=(r_{ij})_{n\times n}$ *and* $R^{-1}=(\tilde{r}_{ij})_{n\times n}$ *are given by*

$$r_{ij}=q_{i-1}[\lambda_1;\ldots;\lambda_j],\qquad \tilde{r}_{ij}=p_{j-1}[\lambda_i;\ldots;\lambda_n], \tag{3.11}$$

where $q_{i-1}(\lambda)=\lambda^{i-1}$, *and* $p_{j-1}(\lambda)$ *is the* $(j-1)$*st adjoint of* $p(\lambda)$.

This corollary gives us [taking $f(\lambda)\equiv\lambda$] the representation of a similarity transformation matrix between C and the bidiagonal matrix $[\Lambda]$. Combining it with Theorem 3.1, we obtain yet another corollary representing a similarity transformation matrix that reduces C to a direct sum of lower-order companion matrices.

COROLLARY 3.3. *Suppose the notation and assumptions of Theorem* 3.1. *Let* $C^{(k)}$ *be the companion matrix of*

$$r_k(\lambda)=\prod_{\nu=1}^{n_k}\left(\lambda-\lambda_\nu^{(k)}\right),$$

and denote by $r_{k,j}(\lambda)$ *the* j*th adjoint polynomial of* $r_k(\lambda)$ $(k=1,\ldots,s,$ $j=0,1,\ldots,n_k-1)$. *Then*

$$C=T(C^{(1)}\oplus\cdots\oplus C^{(s)})T^{-1}, \tag{3.12}$$

where

$$T=[T^{(1)},\ldots,T^{(s)}],\qquad T^{-1}=\begin{bmatrix}\tilde{T}^{(1)}\\ \vdots\\ \tilde{T}^{(s)}\end{bmatrix}, \tag{3.13a}$$

$$T^{(k)}=\left(\psi_{k;ij}\left[\lambda_1^{(k)};\ldots;\lambda_{n_k}^{(k)}\right]\right)_{n\times n_k},\quad \psi_{k;ij}(\lambda)=\lambda^{i-1}r_{k,j-1}(\lambda),$$

$$\tilde{T}^{(k)}=\left(\tilde{\psi}_{k;ij}\left[\lambda_1^{(k)};\ldots;\lambda_{n_k}^{(k)}\right]\right)_{n_k\times n},\quad \tilde{\psi}_{k;ij}(\lambda)=\lambda^{i-1}p_{k,j-1}(\lambda),$$

$$k=1,\ldots,s. \tag{3.13b}$$

Proof. Let $f(\lambda)\equiv\lambda$. From Corollary 3.2 we get

$$[\Lambda^{(k)}]=(R^{(k)})^{-1}C^{(k)}R^{(k)},\qquad k=1,\dots,s, \tag{3.14}$$

where $R^{(k)}$ and its inverse are given according to (3.11). Substituting (3.14) into (3.5)–(3.6), we find that the decomposition (3.12) holds with

$$T^{(k)}=Q^{(k)}(R^{(k)})^{-1}\text{ and }\tilde{T}^{(k)}=R^{(k)}\tilde{Q}^{(k)},\qquad k=1,\dots,s. \tag{3.15}$$

Writing down the matrix products in (3.15) elementwise, and using the divided-difference representations in (3.6) and (3.11), the Leibniz product rule implies the representation in (3.13). ■

4. SIMILARITY TRANSFORMATION OF $f(C)$ TO COMPANION FORM

In the study of polynomial root location the following problem arises [15, 17]: find a matrix that transforms the linear fractional function $\varphi(C)=(\alpha C+\beta I)(\gamma C+\delta I)^{-1}$ of the companion matrix C by similarity to companion form. Such a transformation matrix—with the additional property that it depends only on $\alpha,\beta,\gamma,\delta$ and the order n of C, but not on the elements of C—has been found by Shane and Barnett [15]. A complete proof of their result based on a characterization of companion matrices was given by Young [17].

In this section the above problem is solved for a general function $f(\lambda)$ by giving a (constructive) characterization of all similarity transformation matrices that reduce $f(C)$ to companion form. Since in case of a general function $f(\lambda)$ such a transformation is not always possible [inspect, e.g., the case $f(\lambda)\equiv 0$, $n>1$], we first consider the question of when is $f(C)$ similar to a companion matrix.

If C is the companion matrix of the polynomial

$$p(\lambda)=\prod_{i=1}^{n}(\lambda-\lambda_i), \tag{4.1}$$

then the eigenvalues of $f(C)$ are $f(\lambda_1),\dots,f(\lambda_n)$, so the companion matrix similar to $f(C)$ may only be the companion matrix $C_{(f)}$ of the polynomial

$$p_{(f)}(\lambda)=\prod_{i=1}^{n}[\lambda-f(\lambda_i)]. \tag{4.2}$$

The next theorem gives some criteria for the similarity of $f(C)$ and $C_{(f)}$. Criterion (1) was suggested by R. E. Hartwig.

THEOREM 4.1. *Let C and $C_{(f)}$ be the companion matrices of the polynomials $p(\lambda)$ and $p_{(f)}(\lambda)$ defined by* (4.1) *and* (4.2) *(respectively). The following four conditions are equivalent:*

(0) *$f(C)$ is similar to $C_{(f)}$;*
(1) *$f(C)$ is nonderogatory;*
(2) *there is a function $\tilde{f}(\lambda)$ such that $\tilde{f}(f(C)) = C$;*
(3) *the eigenvalues of C satisfy the condition*

$$f(\lambda_i) \neq f(\lambda_j) \qquad \text{if} \quad \lambda_i \neq \lambda_j,$$

$$f'(\lambda_i) \neq 0 \qquad \text{if} \quad \lambda_i = \lambda_j \text{ and } i \neq j$$

for $i, j = 1, \ldots, n$.

Proof. We prove the cycle of implications $(0) \Rightarrow (1) \Rightarrow \cdots \Rightarrow (0)$. $(0) \Rightarrow (1)$ is trivial, since the companion matrix $C_{(f)}$ is nonderogatory. $(1) \Rightarrow (2)$: As C and $f(C)$ commute, $f(C)$ nonderogatory implies (by Corollary 1 on p. 222 of [7]) that there exists a polynomial $\tilde{f}(\lambda)$ such that $C = \tilde{f}(f(C))$. $(2) \Rightarrow (3)$: If condition (3) is not satisfied, then $f(C)$—and consequently $\tilde{f}(f(C))$—either has a smaller number of different eigenvalues than C (if $f(\lambda_i) = f(\lambda_j)$ when $\lambda_i \neq \lambda_j$) or is derogatory (if $f'(\lambda_i) = 0$ for a multiple eigenvalue λ_i: cf. [7, p. 158, Theorem 9]). Both possibilities contradict (2). $(3) \Rightarrow (0)$ follows immediately from Theorem 9 on p. 158 of [7]. ∎

Now we are able to give the representation of all similarity-transformation matrices (and that of their inverses) that reduce $f(C)$ to companion form.

THEOREM 4.2. *Let the companion matrices C and $C_{(f)}$, the polynomials $p(\lambda)$ and $p_{(f)}(\lambda)$, and the function $\tilde{f}(\lambda)$ be given as in Theorem 4.1, assuming $f(C)$ nonderogatory. Then any nonsingular matrix U satisfying*

$$Uf(C)U^{-1} = C_{(f)} \tag{4.3}$$

is characterized by the relation

$$U = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_1^T f(C) \\ \vdots \\ \mathbf{e}_1^T [f(C)]^{n-1} \end{bmatrix} g(C) \tag{4.4}$$

with a function $g(\lambda)$ *such that* $g(C)$ *is nonsingular (i.e.,* $g(\lambda_i) \neq 0$, $i = 1, \dots, n$*). Moreover, the inverse of the matrix* U *given by* (4.4) *can be represented as*

$$U^{-1} = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_1^T \tilde{f}(C_{(f)}) \\ \vdots \\ \mathbf{e}_1^T [\tilde{f}(C_{(f)})]^{n-1} \end{bmatrix} [g(\tilde{f}(C_{(f)}))]^{-1}. \tag{4.5}$$

Proof. Applying Lemma 3.2 (folklore lemma [9]) to the equation $Uf(C) = C_{(f)}U$, it follows that

$$U^T = \left[\mathbf{y}, (f(C))^T \mathbf{y}, \dots, ((f(C))^T)^{n-1} \mathbf{y}\right] \tag{4.6}$$

with some n-vector $\mathbf{y}$. Defining $g(\lambda) = y_1 + y_2\lambda + \cdots + y_n\lambda^{n-1}$ and using Lemma 2.1 we have

$$\mathbf{y}^T = \sum_{i=1}^{n} y_i \mathbf{e}_i^T = \mathbf{e}_1^T \sum_{i=1}^{n} y_i C^{i-1} = \mathbf{e}_1^T g(C),$$

which, combined with (4.6), yields the desired representation (4.4). Also, part (c) of the folklore lemma implies U is nonsingular if and only if $g(\lambda)$ and $p(\lambda)$ have no roots in common, that is, $g(\lambda_i) \neq 0$, $i = 1, \dots, n$.

To obtain the inverse U^{-1} we utilize the similarity relation

$$UCU^{-1} = \tilde{f}(C_{(f)}), \tag{4.7}$$

which follows from $\tilde{f}(f(C)) = C$. Using (4.7) and observing that $\mathbf{e}_1^T U =$

$\mathbf{e}_1^T g(C)$, we can determine the first row of U^{-1}:

$$\mathbf{e}_1^T U^{-1} = \mathbf{e}_1^T U [g(C)]^{-1} U^{-1} = \mathbf{e}_1^T \left[g\left(\tilde{f}(C_{(f)}) \right) \right]^{-1}.$$

Now an application of the folklore lemma to the equation $CU^{-1} = U^{-1} \tilde{f}(C_{(f)})$ immediately yields (4.5). ■

Using Theorem 2.1, we can directly obtain the divided-difference representations of the entries of the transformation matrix U and its inverse.

Corollary 4.1. *The transformation matrix* $U = (u_{ij})_{n \times n}$ *and its inverse* $U^{-1} = (\tilde{u}_{ij})_{n \times n}$ *given by* (4.4)–(4.5) *in Theorem* 4.2 *have the divided-difference representations*

$$u_{ij} = \omega_{ij}[\lambda_1; \ldots ; \lambda_n], \qquad \omega_{ij}(\lambda) = [f(\lambda)]^{i-1} g(\lambda) p_{j-1}(\lambda) \quad (4.8)$$

and

$$\tilde{u}_{ij} = \tilde{\omega}_{ij}[f(\lambda_1); \ldots ; f(\lambda_n)], \qquad \tilde{\omega}_{ij}(\lambda) = \frac{[\tilde{f}(\lambda)]^{i-1} p_{(f)j-1}(\lambda)}{g(\tilde{f}(\lambda))} \quad (4.9)$$

where $p_{j-1}(\lambda)$ *and* $p_{(f)j-1}(\lambda)$ *denote the* $(j-1)$*st adjoint polynomials of* $p(\lambda)$ *and* $p_{(f)}(\lambda)$, *respectively.*

Example (linear fractional functions). As an example, let us consider the case of a linear fractional function

$$f(\lambda) = \frac{\alpha\lambda + \beta}{\gamma\lambda + \delta}$$

investigated in the papers [15, 17]. Here the trivial case $f(\lambda) \equiv \text{const.}$ is excluded by the condition $D = \alpha\delta - \beta\gamma \neq 0$. Then $f(C)$ exists whenever the eigenvalues λ_i of the companion matrix C are not poles of $f(\lambda)$, that is, if $\gamma\lambda_i + \delta \neq 0$, $i = 1, \ldots, n$. In this case, $f(C)$ is always similar to $C_{(f)}$, as condition (3) (e.g.) of Theorem 4.1 is fulfilled [$f(\lambda)$ is one-to-one and $f'(\lambda) \neq 0$]. For the application of Theorem 4.2 and Corollary 4.1 to determine a similarity transformation matrix $U = (u_{ij})_{n \times n}$ we can choose $g(\lambda) = (\gamma\lambda + \delta)^{n-1}$, since $g(C)$ is nonsingular whenever $f(C)$ exists. Then we have

$u_{ij} = \omega_{ij}[\lambda_1;\dots;\lambda_n]$ with

$$\begin{aligned}\omega_{ij}(\lambda) &= \left(\frac{\alpha\lambda+\beta}{\gamma\lambda+\delta}\right)^{i-1}(\gamma\lambda+\delta)^{n-1}p_{j-1}(\lambda)\\ &= (\alpha\lambda+\beta)^{i-1}(\gamma\lambda+\delta)^{n-i}p_{j-1}(\lambda)\\ &= \sum_{k=1}^{n} m_{ik}\lambda^{k-1}p_{j-1}(\lambda), \qquad i,j=1,\dots,n,\end{aligned}$$

where m_{ik} denotes the coefficient of λ^{k-1} in the expanded form of the polynomial $(\alpha\lambda+\beta)^{i-1}(\gamma\lambda+\delta)^{n-i}$. However, by Corollary 2.1 we have

$$u_{ij} = \omega_{ij}[\lambda_1;\dots;\lambda_n] = \sum_{k=1}^{n} m_{ik}\delta_{kj} = m_{ij}, \qquad i,j=1,\dots,n,$$

independent of the eigenvalues of C, and this is just the result of Shane and Barnett [15].

Using Theorem 4.2, we can also determine the inverse of the Shane-Barnett transformation matrix $M=(m_{ij})_{n\times n}$. For, we can choose $\tilde{f}(\lambda)$ as the inverse function

$$\tilde{f}(\lambda) = \frac{\delta\lambda-\beta}{\alpha-\gamma\lambda} = \frac{\tilde{\alpha}\lambda+\tilde{\beta}}{\tilde{\gamma}\lambda+\tilde{\delta}},$$

with $\tilde{\alpha}=\delta/D$, $\tilde{\beta}=-\beta/D$, $\tilde{\gamma}=-\gamma/D$, $\tilde{\delta}=\alpha/D$, and $D=\alpha\delta-\beta\gamma$. It is clear that $\tilde{f}(\lambda)$ satisfies $\tilde{f}(f(C))=C$ whenever $f(C)$ exists. Also, a simple calculation shows that $[g(\tilde{f}(\lambda))]^{-1}=(\tilde{\gamma}\lambda+\tilde{\delta})^{n-1}$; thus we can proceed completely analogously to the above to obtain $M^{-1}=(\tilde{m}_{ij})_{n\times n}$, where $\tilde{m}_{ij}$ is the coefficient of λ^{j-1} in the expanded form of the polynomial $(\tilde{\alpha}\lambda+\tilde{\beta})^{i-1}(\tilde{\gamma}\lambda+\tilde{\delta})^{n-i}$.

In special cases of a linear fractional function the entries of M and M^{-1} can be given in particularly simple explicit forms. For the linear functions $f(\lambda)=\lambda-a$ and $f(\lambda)=\lambda/b$ the explicit forms were derived (in a direct way) by Kammler [10], who used the similarity transformation of $b^{-1}(C-aI)$ to companion form in preprocessing a companion matrix for the computation of the matrix exponential $\exp(tC)$. In the case of the reciprocal function $f(\lambda)=1/\lambda$ we have $\tilde{f}(\lambda)=1/\lambda$, as well as

$$\sum_{j=1}^{n} m_{ij}\lambda^{j-1} = \sum_{j=1}^{n} \tilde{m}_{ij}\lambda^{j-1} = (0\cdot\lambda+1)^{i-1}(1\cdot\lambda+0)^{n-i} = \lambda^{n-i};$$

hence

$$M = M^{-1} = \left(\delta_{j-1,n-i}\right)_{n\times n},$$

the "reverse ordering" permutation matrix.

5. CONNECTION BETWEEN FUNCTIONS OF GENERAL AND COMPANION MATRICES

In this short section we prove a slightly generalized version of a result of Stafney [16], complemented with a dual relationship in terms of adjoint polynomials.

THEOREM 5.1. *Let A be an $n\times n$ matrix, $p(\lambda)=c_0+c_1\lambda+\cdots+c_{n-1}\lambda^{n-1}+\lambda^n$ an annihilating polynomial of A, and C the companion matrix of $p(\lambda)$, and suppose that $f(A)$ and $f(C)=(f_{ij})_{n\times n}$ both exist. Then the dual pair of relations*

$$f(A)=\sum_{j=1}^{n} f_{1j}A^{j-1} \tag{5.1}$$

and

$$f(A)=\sum_{i=1}^{n} f_{in}p_{i-1}(A) \tag{5.2}$$

holds, where $p_{i-1}(\lambda)$ is the $(i-1)$st adjoint polynomial of $p(\lambda)$.

Proof. The relation $p(A)=0$ implies, according to the power-shift condition of Hartwig [8], that for any polynomial $f(\lambda)$

$$[f(C)\otimes I]Q_A=Q_Af(A) \tag{5.3}$$

where $\otimes$ denotes the Kronecker product, I is $m\times m$, and Q_A is $m^2\times m$ given by

$$Q_A=\begin{bmatrix} I\\ A\\ \vdots\\ A^{m-1}\end{bmatrix}.$$

To verify (5.3) for a general function $f(\lambda)$, choose $h(\lambda)$ as a polynomial that agrees with $f(\lambda)$ on the spectrum of C, so that $f(C) = h(C)$. Since the spectrum of A is contained in that of C $[p(A) = 0]$, we have also $f(A) = h(A)$, proving (5.3) in general. Equating the first n rows in (5.3) yields (5.1).

The dual relation (5.2) can be derived from (5.1) using Lemma 2.1 and the definition of adjoint polynomials, as follows:

$$f(A) = \sum_{j=1}^{n} \mathbf{e}_1^T f(C)\mathbf{e}_j A^{j-1} = \sum_{j=1}^{n} \mathbf{e}_1^T f(C) p_{j-1}(C)\mathbf{e}_n A^{j-1}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n+1-j} c_{i-1+j} \mathbf{e}_1^T C^{i-1} f(C)\mathbf{e}_n A^{j-1}$$

$$= \sum_{i=1}^{n} \left(\mathbf{e}_i^T f(C)\mathbf{e}_n\right)\left(\sum_{j=1}^{n+1-i} c_{i+j-1} A^{j-1}\right) = \sum_{i=1}^{n} f_{in} p_{j-1}(A). \quad \blacksquare$$

Remark 5.1. Stafney [16] proved the relation (5.1) of Theorem 5.1 in the case $f(\lambda)$ a polynomial and $p(\lambda)$ the minimal polynomial of A. For the special case of the matrix exponential $\exp(tA)$ a number of proofs to (5.1) have been given, using spectral decomposition, differential equations, etc. In particular, Kolodner [11] proved this case in an abstract setting, using differential equations. His proof was adapted by this author [5] for the computation of $\exp(tA)\mathbf{b}$ in terms of the Krylov vectors $A^{j-1}\mathbf{b}$ and the companion matrix of the minimal polynomial of the vector $\mathbf{b}$. In this regard it may be mentioned that Theorem 5.1 with its proof remains valid if the expansion of $f(A)B$ in terms of $A^{j-1}B$ or $p_{i-1}(A)B$ is considered, where B is a rectangular matrix and $p(\lambda)$ a polynomial satisfying $p(A)B = 0$ (e.g., the minimal polynomial of the column space of B).

The author is deeply indebted to Professor Hartwig for his many helpful suggestions, which resulted in the simplification of almost all proofs in the paper.

REFERENCES

1 I. S. Berezin and N. P. Zhidkov, *Computing Methods*, Vol. 1, Pergamon, Oxford, 1965.
2 L. Brand, The companion matrix and its properties, *Amer. Math. Monthly* 71:629–634 (1964).

3 E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable*, Clarendon, Oxford, 1935.
4 F. G. Csáki, Some notes on the inversion of confluent Vandermonde matrices, *IEEE Trans. Automat. Control* AC-20:154–157 (1975).
5 J. Eller, Numerical determination of the matrix exponential and its biological application (in Hungarian), Univ. Doct. Dissert., József Attila Univ., Szeged, 1982.
6 J. W. Evans, W. B. Gragg, and R. J. LeVegue, On least squares exponential sum approximation with positive coefficients, *Math. Comp.* 34:203–211 (1980).
7 F. R. Gantmacher, *The Theory of Matrices*, Vol. 1, Chelsea, New York, 1959.
8 R. E. Hartwig, Resultants and the solution of $AX - XB = -C$, *SIAM J. Appl. Math.* 23:104–117 (1972).
9 R. E. Hartwig, unpublished manuscript.
10 D. W. Kammler, Numerical evaluation of $\exp(tA)$ when A is a companion matrix, *SIAM J. Numer. Anal.* 15:1077–1102 (1978).
11 I. I. Kolodner, On $\exp(tA)$ with A satisfying a polynomial, *J. Math. Anal. Appl.* 52:514–524 (1975).
12 C. Moler and C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, *SIAM Rev.* 20:801–836 (1978).
13 G. Opitz, Steigungsmatrizen, *Z. Angew. Math. Mech.* 44:T52—T54 (1964).
14 R. F. Rinehart, The equivalence of definitions of a matric function, *Amer. Math. Monthly* 62:395–414 (1955).
15 B. A. Shane and S. Barnett, On the bilinear transformation of companion matrices, *Linear Algebra Appl.* 9:175–184 (1974).
16 J. D. Stafney, Functions of a matrix and their norms, *Linear Algebra Appl.* 20:87–94 (1978).
17 N. J. Young, Linear fractional transforms of companion matrices, *Glasgow Math. J.* 20:129–132 (1979).

Received 23 September 1982; revised 7 January 1987